

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/31499 A2

(51) International Patent Classification⁷: G06F 17/30

(21) International Application Number: PCT/GB00/03503

(22) International Filing Date:
11 September 2000 (11.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9925400.5 27 October 1999 (27.10.1999) GB

(71) Applicant and

(72) Inventor: HARPAZ, Yehouda [GB/GB]; 129 Corrie
Road, Cambridge, Cambridgeshire CB1 3QQ (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,

DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished
upon receipt of that report.

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: CONSTRAINED DOCUMENT SERVER

(57) Abstract: A constraint document server, comprising a database of documents and constraints for each document, and software to search the document, to extract pieces of texts out of them, and to limit the amount of text that is being extracted according to the constraints. The software includes a *searcher* component, which searches the documents and identifies potential matching documents, and *extractor/filter* component, which extract text from the potential matching documents, and checks that the extracted text does not exceed the constraints.

WO 01/31499 A2